

**UTILISATION D'UN MODELE DE REPONSE A L'ITEM POUR LA  
VALIDATION D'UNE ECHELLE DES ATTITUDES PROFESSIONNELLES  
DES MEDECINS GENERALISTES**

**Etienne Le Bihan<sup>1</sup>, Cédric Baumann<sup>23</sup>, François Alla<sup>23</sup>, Michèle Baumann<sup>1</sup>, (1)**

Integrative research unit on Social and Individual DEvelopment (INSIDE),  
Université du Luxembourg, (2) Service d'Epidémiologie et d'Evaluation Cliniques,  
CEC – INSERM, Centre Hospitalier Universitaire de Nancy, (3) Ecole de Santé  
Publique, EA 4003, Faculté de Médecine, Université de Nancy, France.

**Mots-clés :** Echelle d'Attitudes Professionnelles des Médecins Généralistes,  
Modèle de Réponse à l'Item, Fonctionnement Différentiel d'Item.

**Résumé :**

Nous avons développé une échelle des attitudes professionnelles des médecins généralistes destinée à évaluer le niveau de communication thérapeutique dispensé durant les consultations. La construction de l'échelle a abouti à un questionnaire constitué de 15 items à 10 modalités de réponse allant de "Jamais" à "Toujours". Après avoir étudié les propriétés métriques de l'échelle selon une méthodologie classique, nous avons employé un modèle de réponse à l'item (le Generalized Partial Credit Model, GPCM) pour explorer un éventuel fonctionnement différentiel d'item (FDI) selon le genre des individus. Le modèle a été construit dans un cadre bayésien en utilisant un algorithme MCMC. La détection du FDI est basée sur le test de rapport de vraisemblance, en comparant le modèle de référence où tous les paramètres associés aux items sont communs aux hommes et aux femmes, à un modèle incluant deux ensembles de paramètres distincts selon le genre des individus pour un item particulier. La démarche a été appliquée aux 338 observations de l'échantillon de validation ne comportant pas de données manquantes. Parmi les 15 items de l'échelle, seul l'item 6 ("Mon médecin m'informe sur les effets secondaires des médicaments") présente un FDI significatif au niveau de signification de 5% ( $p = 0.016$ ). Cet item est plus discriminant chez les hommes. De plus, il semble que, à un niveau de communication thérapeutique égal, les hommes ont une probabilité supérieure aux femmes de répondre aux catégories les plus élevées. Finalement, nous avons décidé de ne pas écarter cet item car 1) le FDI est modéré (si nous ajustons le test pour les comparaisons multiples, celui-ci n'est plus significatif au niveau 5%) et 2) nous n'avons aucune interprétation qualitative de cette différence.

275 mots.

---

**Introduction**

**Contexte**

Le développement d'une échelle d'attitudes des médecins généralistes a pour objectif de mesurer le niveau de communication thérapeutique dispensé par les médecins généralistes durant les consultations.

Selon l'Organisation Mondiale de la Santé, l'éducateur thérapeutique est "celui qui permet aux patients et à leur famille d'acquérir et de conserver les capacités et les compétences qui les aident à vivre de manière optimale leur vie avec leur maladie" (Organisation Mondiale de la Santé [OMS], 1998).

Dans le but d'améliorer l'état de santé des populations et de limiter les dépenses de santé, ce concept fait l'objet de nombreux programmes de politique de santé publique pour lesquels l'échelle des attitudes professionnelles des médecins généralistes à l'égard de la communication thérapeutique (GPPATC = General Partitionner Professional Attitudes towards Therapeutic Communication) a pour objectif de fournir un outil d'évaluation.

### Construction des items du questionnaire et développement de l'échelle \*

Le développement de l'échelle a fait l'objet d'une précédente publication (Baumann, Baumann, Aubry, Alla, 2005). Il s'est appuyé sur le recueil du discours de 40 usagers lorrains d'un centre de médecine préventive, 21 médecins généralistes et 22 pharmaciens d'officine de l'Est de la France. L'analyse de contenu a permis de retenir les items candidats parmi lesquels des experts ont sélectionnés les items génériques du questionnaire. Au final 15 items ont été retenus dans le questionnaire (tableau 1). A chacun des items, dix modalités de réponse sont proposées variant de "jamais" à "toujours".

**Tableau 1** - Les items de l'échelle des attitudes professionnelles des médecins généralistes.

No	Item
1	Mon médecin prend le temps de m'écouter
2	Mon médecin fait le nécessaire pour gagner ma confiance
3	Mon médecin m'explique à quoi sert le traitement
4	Mon médecin tient compte de mes préférences pour la prescription (cachet, gouttes etc.)
5	Mon médecin me donne l'impression qu'il me respecte
6	Mon médecin m'informe sur les effets secondaires des médicaments
7	Mon médecin insiste sur les médicaments importants
8	Mon médecin parle avec moi des difficultés que j'ai à suivre le traitement
9	Mon médecin m'explique les choses avec des mots simples
10	Mon médecin me propose des nouveaux traitements
11	Mon médecin écrit lisiblement l'ordonnance
12	Mon médecin me laisse poser mes questions
13	Mon médecin me motive pour suivre mon traitement
14	Mon médecin me donne des conseils de prévention (alimentation, activité sportive...)
15	Mon médecin me donne l'impression de connaître son métier

\* Michèle Baumann, assistante professeure en sociologie de la santé, responsable scientifique du programme, a obtenu le soutien financier de la Caisse Nationale d'Assurance Maladie, Paris, France.

### Validation « classique »

#### Echantillon

L'échantillon pour la validation statistique du questionnaire se compose de 393 sujets recrutés dans un Centre Médecine Préventive en Lorraine en 2003. Les sujets ont été invités à remplir un deuxième questionnaire 15 jours après la première passation. L'âge moyen des sujets de l'échantillon était de 46.8 ans. La parité hommes / femmes est quasi-parfaite (50,4% d'hommes versus 49,6% de femmes).

#### Résultats

Les résultats de la validation classique montrent que l'échelle est unidimensionnelle puisque le premier facteur de l'Analyse en Composantes Principales représente 49.8% de l'inertie totale (autres facteurs : 8.8% à 1.2%). Tous les items sont bien représentés sur ce facteur. La cohérence interne est élevée (alpha de Cronbach = 0.92) et reproductibilité satisfaisante (Coefficient de corrélation intra-classe = 0.74 [0.66 ; 0.82]). La figure 1 montre cependant un effet seuil important pour certains items, ce qui suggère que l'échelle est peu discriminante pour des valeurs d'attitude élevées.

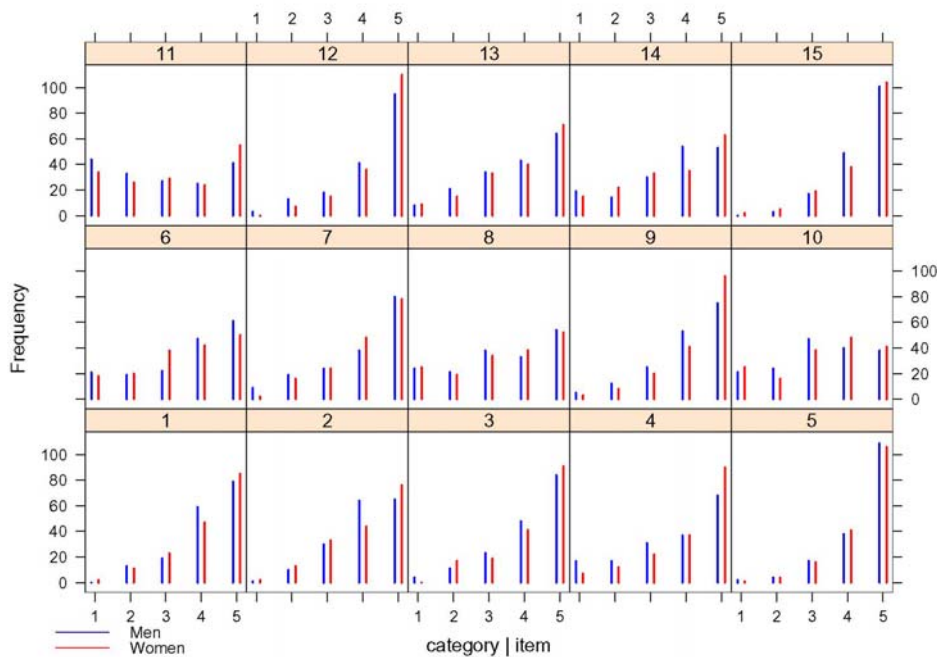


Figure 1 - Distribution des fréquences de réponse aux items.

#### Recherche d'un Fonctionnement Différentiel d'Item Hommes/Femmes

### Définition

Un Fonctionnement Différentiel d'Item (FDI) se produit lorsque, pour un niveau d'attitude donné, la réponse attendue est différente selon le groupe d'appartenance de la personne qui a répondu au questionnaire (Penfield & Lam, 2000).

### Le Generalized Partial Credit Model

Afin de tester la présence éventuelle d'un FDI, nous nous sommes appuyés sur un modèle de réponse à l'item, le Generalized Partial Credit Model (Muraki, 1992). Selon l'équation 1, le modèle associe la probabilité pour un individu  $i$  de répondre à la catégorie  $k$  de l'item  $j$  sachant qu'il a répondu à la catégorie  $k$  ou  $k-1$ , à 3 paramètres :

- $\theta_i$  : mesure de l'attitude thérapeutique du sujet  $i$ ,
- $b_{jk}$  : « difficulté » pour passer de la catégorie  $k-1$  à la catégorie  $k$ ,
- $a_j$  : discrimination de l'item  $j$ .

$$\text{Logit}(P_{ijk|k-1,k}) = a_j (\theta_i - b_{jk}) \quad (1)$$

### Démarche

En supposant que le modèle est correct, les paramètres  $\{a_j, b_{jk}\}$  résume totalement les propriétés métriques de l'item  $j$ . Ainsi, le FDI d'un item particulier  $j$  se traduit par le fait que les paramètres  $\{a_j, b_{jk}\}$  diffèrent selon le groupe auquel l'item est proposé. Notre problématique se résume alors à tester l'égalité des 2 ensembles de paramètres  $\{a_j^H, b_{jk}^H\}$  et  $\{a_j^F, b_{jk}^F\}$ , ( $H$  = Hommes,  $F$  = Femmes).

Envisageons les deux modèles GPCM suivants :

1.  $\{a_j, b_{jk}\}$  pour tout  $j$  : tous les paramètres sont communs aux hommes et aux femmes,
2.  $\{a_j^H, b_{jk}^H, a_j^F, b_{jk}^F\}$  ;  $\{a_j, b_{jk}\}$ , pour tout  $j \neq j'$  : les paramètres de l'item  $j'$  sont supposés différents entre les hommes et les femmes, les autres paramètres sont communs aux hommes et aux femmes.

Le modèle 1 étant un cas particulier du modèle 2, ces deux modèles sont emboîtés. Nous pouvons donc les comparer par le test du rapport du maximum de vraisemblance : sous l'hypothèse que le modèle 1 est correct, on démontre que

$$-2\{\text{Log}(L_1) - \text{Log}(L_2)\} \sim \chi^2_{(n_2 - n_1)} \quad (2)$$

où  $L_1$  et  $L_2$  désignent respectivement les vraisemblances des modèles 1 et 2,  $n_1$  et  $n_2$  étant les nombres de paramètres respectifs des modèles 1 et 2.

### Mise en oeuvre

La mise en oeuvre s'appuie sur les algorithmes MCMC dans un cadre bayésien. Tous les modèles ont été ajustés grâce au logiciel WinBUGS (Spiegelhalter, Thomas, Best, Lunn, 2003). Le GPCM étant sur-paramétré, nous avons imposé aux paramètres associés à la première catégorie d'être nuls, quelque soit l'item. De même, la contrainte classique selon laquelle la distribution des  $\theta_i$  est centrée et de variance 1 a

été retranscrite en posant une loi a priori centrée réduite pour ces paramètres. Notons que cela n'est pas strictement équivalent, ce qui a pour conséquence que les estimations des paramètres ne sont pas immédiatement comparables à ceux fournis par les logiciels classiques. Néanmoins, les estimations sont liées entre elles par des relations linéaires dans lesquelles interviennent la moyenne et la variance de la distribution estimée des  $\theta_i$ .

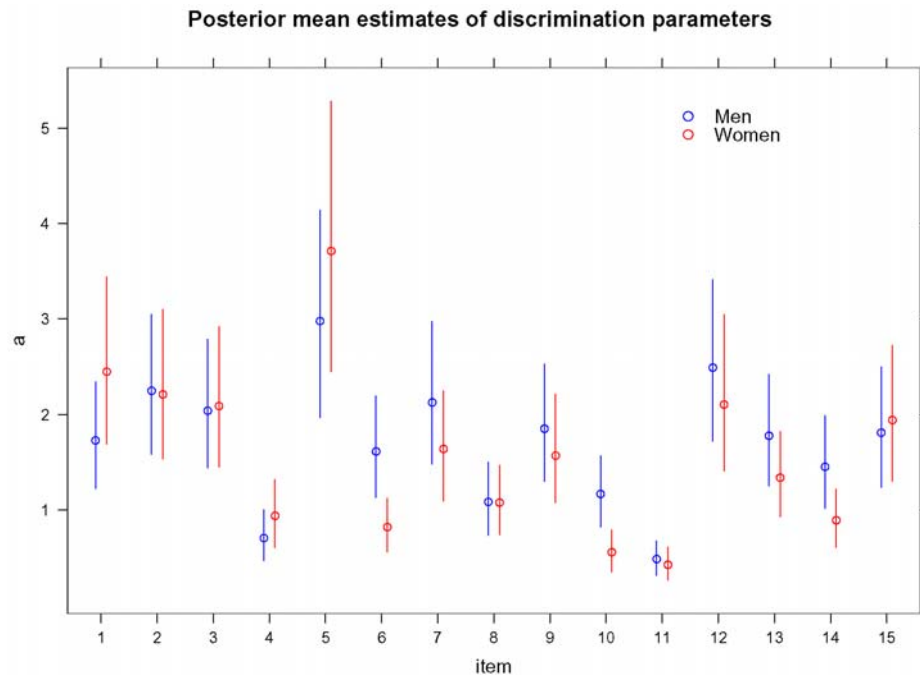
Nous avons posé des lois a priori non-informatives sur les autres paramètres : lois normales centrées de variance 100 et limitées aux valeurs positives pour les paramètres de discrimination, lois normales centrées de variance 100 pour les paramètres de  $c_{jk}$  (la paramétrisation du modèle a été modifiée en posant  $c_{jk} = b_{jk} a_j$  afin d'améliorer la convergence des algorithmes MCMC).

### Données

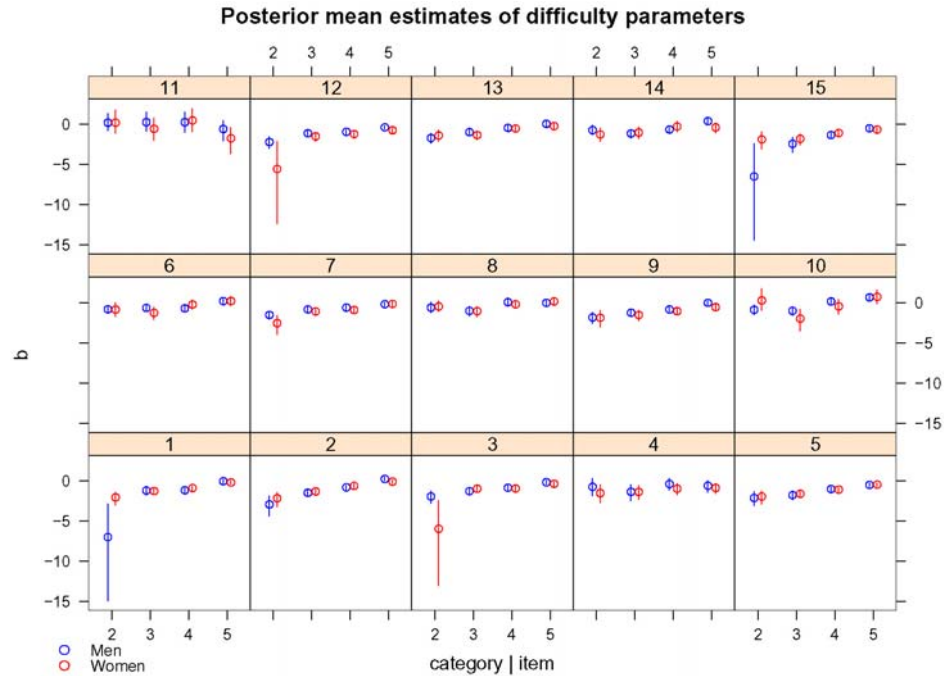
Les modèles ont été ajustés sur les observations recueillies durant la première passation. Les observations comportant au moins une donnée manquante ont été écartées. Au final l'échantillon est composé de 338 observations.

Pour des raisons techniques, les réponses appartenant à des catégories adjacentes ont été regroupées 2 par 2.

### Résultats



**Figure 2** - Estimations et intervalles de crédibilité des paramètres de discrimination. Ces estimations proviennent de l'ajustement d'un modèle avec deux ensembles distincts de paramètres selon le sexe des individus pour chacun des items.



**Figure 3** - Estimations et intervalles de crédibilité des paramètres de difficulté. Ces estimations proviennent de l'ajustement d'un modèle avec deux ensembles distincts de paramètres selon le sexe des individus pour chacun des items.

Les figures 3 et 4 montrent respectivement les estimations des paramètres de discrimination et de difficulté des items lorsque les ensembles de paramètres sont distincts chez les hommes et chez les femmes pour tous les items. On constate que les paramètres de discrimination sont proches chez les hommes et les femmes pour la plupart des items, excepté les items 6 et 10 pour lesquels les intervalles de crédibilité ne se chevauchent pas. On peut observer sur la figure 4 de très larges intervalles de crédibilité associés à des estimations fortement négatives pour la catégorie 2 des items 1, 3, 12 et 15 chez les hommes ou les femmes. Ceci s'explique aisément par le fait qu'aucun individu du sexe correspondant n'ait répondu à la première catégorie de ces items. Aucune différence notable n'est constatée sur les autres estimations des paramètres de difficulté.

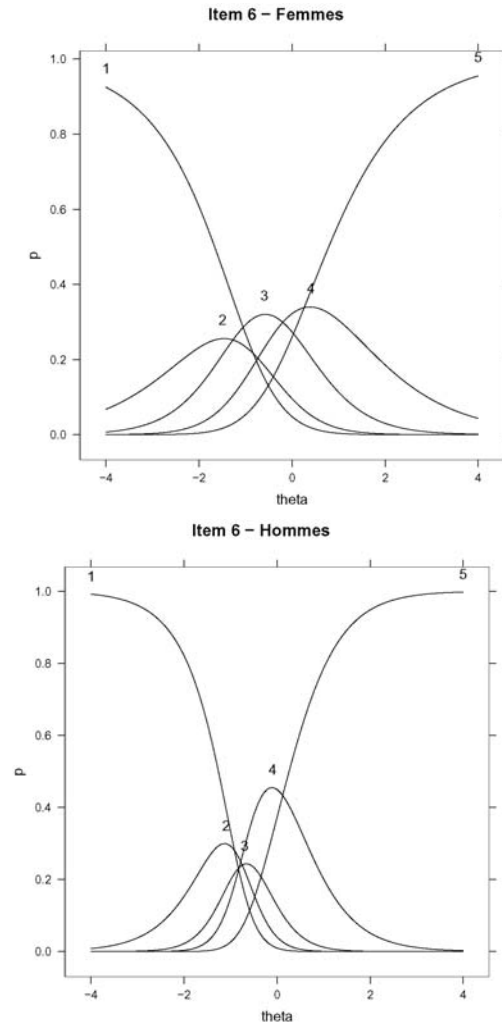
**Tableau 2** - Tests du FDI de chacun des items de l'échelle d'attitudes des médecins généralistes. Chaque ligne correspond à un modèle : la première correspond au modèle (1) où l'ensemble de paramètres associés à chacun des items est commun aux hommes et aux femmes, les suivantes correspondent au modèle (2), j'prenant successivement les valeurs indiquées dans la première colonne "Item". Les valeurs affichées dans la colonne "p" sont les probabilités pour que l'item indiqué en colonne 1 ne présente pas de FDI.

Item	-2LogL	Nombre paramètres	$\chi^2$	ddl	p
	10347.84	75			
1	10341.94	80	5.90	5	0.316
2	10343.63	80	4.21	5	0.520
3	10348.25	80	0.41	5	0.995
4	10341.87	80	5.97	5	0.309
5	10350.71	80	2.87	5	0.721
6	10333.97	80	13.87	5	0.016
7	10348.66	80	0.82	5	0.976
8	10352.85	80	5.01	5	0.414
9	10344.16	80	3.68	5	0.597
10	10341.26	80	6.58	5	0.254
11	10349.97	80	2.13	5	0.831
12	10343.41	80	4.43	5	0.489
13	10350.45	80	2.61	5	0.761
14	10342.15	80	5.69	5	0.338
15	10338.7	80	9.14	5	0.104

Seul le test du FDI de l'item 6 est significatif au niveau de 5% (tableau 2). Il s'agit de l'item "Mon médecin m'informe sur les effets secondaires des médicaments". L'observation des courbes caractéristiques de cet item chez les hommes et les femmes (figure 5) montrent une discrimination plus élevée chez les hommes. Lorsque l'attitude est proche de zéro, on constate aussi que la probabilité de répondre à la catégorie 4 est nettement plus élevée que celle de répondre à la catégorie 3 pour les hommes, alors que ces deux probabilités sont du même ordre chez les femmes.

Finalement, nous avons décidé de ne pas rejeter cet item pour les raisons suivantes :

- pas d'interprétation qualitative du FDI de cet item,
- le FDI est modéré, d'autant plus que nous n'avons pas tenu compte des comparaisons multiples (si on ajuste le test pour les comparaisons multiples par la méthode de Bonferroni, le test n'est plus significatif). On peut donc penser que son influence sur l'échelle est négligeable.



**Figure 4** - Courbes caractéristiques de l'item 6 chez les femmes (à gauche) et chez les hommes (à droite).

### Conclusion

Nous n'avons pas évoqué ici les aspects touchant à la métrique auxquels il faut pourtant être particulièrement attentif lorsque l'on cherche à mettre en évidence un éventuel fonctionnement différentiel d'item. En effet, si l'on suppose que les items ont des propriétés différentes selon les groupes auxquels ils sont soumis, cela revient à chercher à comparer les résultats obtenus par différents groupes à des échelles différentes. Dans notre cas, il est évident que les items de l'échelle présentent peu ou pas de FDI. Les modèles étant ajustés en supposant que tous les items sauf un ne présentent pas de FDI, l'influence de la métrique est assurément négligeable. Lorsque



cela n'est pas le cas, il existe différentes méthodes permettant de lier les échelles entre les groupes (Kim & Cohen, 1992). Citons la démarche de Lord qui consiste à tester la présence de FDI sur chacun des items, puis à écarter les items dont on suspecte qu'ils ont un FDI et à estimer les paramètres  $\theta_i$ , et enfin à recommencer les tests de FDI en imposant que les paramètres associés aux individus soient égaux aux  $\theta_i$  précédemment estimés.

La méthode est simple et efficace. Son principal désavantage est la durée d'ajustement des modèles (plusieurs heures). Sachant qu'il faut ajuster autant de modèles que l'échelle contient d'items plus un, cela rend la méthode pratiquement inutilisable en routine. Par contre, pour peu que l'on vérifie les conditions de convergence du modèle, l'utilisation des algorithmes MCMC assurent une qualité des estimations qui supporte un statut de méthode de référence. Plusieurs extensions sont envisageables :

- La prise en compte des données manquantes par augmentation des données,
- Modèle hiérarchique : les paramètres  $\theta_i$  sont supposés tirés d'une population plus large,
- Prise en compte des données répétées (plusieurs observations étalées dans le temps chez les mêmes personnes) : dans ce cas, on peut imaginer de poser un modèle mixte sur les  $\theta_i$  conjointement au modèle de réponse à l'item.

### Références

Baumann M., Baumann C., Aubry C., Alla F. (2005). Echelle des attitudes professionnelles des médecins généralistes et des pharmaciens d'officine favorisant l'observance thérapeutique. *Revue Médicale de l'Assurance Maladie*, 36(1):23-33.

Kim S.-H., Cohen A. S. (1992). Effects of linking methods on detection of DIF. *Journal of Educational Measurement*, 29 (1):51-66.

Muraki E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, 16:159-176.

Organisation Mondiale de la Santé (1998). *Education thérapeutique du patient. Programmes de formation continue pour professionnels de soins dans le domaine de la prévention des maladies chroniques*. Bureau Régional pour l'Europe. Copenhagen : ed. OMS, 16.

Penfield R. D., Lam T. C. M. (2000). Assessing differential item functioning in performance assessment: Review and recommendations. *Educational Measurement: Issues and Practice*, 19(3):5-15.

Spiegelhalter D. J., Thomas A., Best N. G., Lunn D. (2003). *WinBUGS Version 1.4 User Manual*. Cambridge : MRC Biostatistics Unit.